HUMAN FACTORS IMPACT ON NDE RELIABILITY

J. C. Spanner, Sr.

Pacific Northwest Laboratory
Operated by Battelle Memorial Institute
P.O. Box 999
Richland, WA 99352

INTRODUCTION

Measurements of nondestructive testing (NDT) system reliability in both the nuclear and non-nuclear segments of industry have consistently produced disappointing results [1,2,3]. While the equipment and techniques appear intrinsically capable of the required performance, overall reliability is often so poor that the credibility of the entire inspection process is jeopardized. On the basis of results accumulated during the past decade, we now know that many NDT processes are not sufficiently reliable to meet the current needs of industry; *but we do not know why*! Furthermore, we expect NDT performance under actual field conditions to be no better, and probably worse, than the performance measured under laboratory conditions. *Significantly, nondestructive testing/inservice inspection (NDT/ISI) performance under actual field conditions remains an important unknown within the nuclear industry.*

Research (under both laboratory and field conditions) is needed to identify, characterize, and quantify the human performance factors and their influence on the overall effectiveness of NDT/ISI. After the important human performance factors have been identified and evaluated, they should be prioritized with respect to relative influence and correctability potential (i.e., field implementation potential). Finally, recommendations should be developed toward improving the human performance factors that adversely affect the reliability of NDT/ISI.

This paper describes a feasibility study that was conducted as Phase 1 of this overall program [4,5]. The objectives of this study were as follows:

- Identify and characterize the human factors aspects of the ultrasonic testing/inservice inspection (UT/ISI) process,
- Develop a model for the UT/ISI man-machine system,
- Examine methods for measuring and analyzing NDT performance, and
- Acquire and evaluate human factors data during a mini-round robin (MRR) test of UT effectiveness in detecting intergranular stress corrosion cracking (IGSCC) in piping specimens.

HUMAN FACTORS ASPECTS OF THE ULTRASONIC TESTING PROCESS

The human factors technology is a multidisciplinary field that utilizes the specialized knowledge available from engineering, experimental psychology, biology, sociology, statistics, and operations research. As shown in Fig. 1, human factors specialists consider the basic human variables as well as the interfaces between the human, the machine, and the environment under which the human-machine system must operate. Human capabilities and limitations, the manner by which humans process and respond to information, and the effect of the environment on overall system performance are important considerations. Many examples where human factors engineering has been successfully applied are documented in the literature [6,7].

A man-machine model was developed to depict the semi-automated NDT system that performs the sensing, information processing, decision making, and action functions (Fig. 2). Training, education, and management/supervisory influence are inputs to this model; and signal interpretation is treated as an information processing, pattern recognition, and decision making process.

While all UT/ISI activities must be performed properly, data acquisition and signal interpretation are key subtasks. Flaw detection and flaw sizing are two separate, and distinctly different, tasks that may be performed by different personnel. Signal interpretation is a human information processing function that includes subjective as well as objective data analysis. Flaw sizing, on the other hand, is generally performed using one or more prescriptive techniques.

MEASUREMENT OF UT RELIABILITY

Three commonly used methods for measuring accuracy were evaluated and discarded when they were found to be unsuitable for quantifying overall UT/ISI system performance. These were: 1) probability of detection, 2) sensitivity, and 3) specificity. We have determined that the ROC analysis method is superior to the other three methods for measuring and describing technician proficiency and NDT system reliability (e.g., during round robin or performance validation tests).

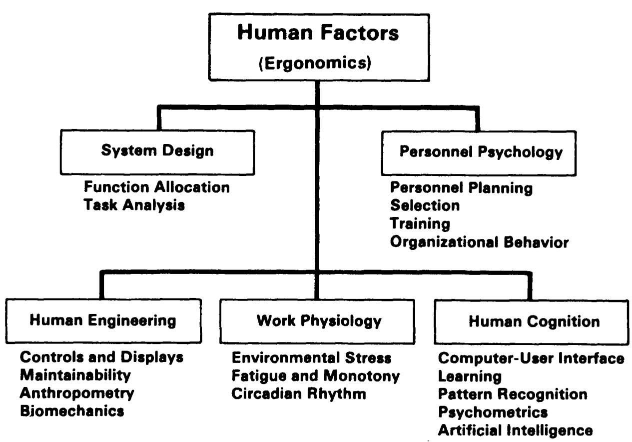

Fig. 1. Five major areas of specialization within the human factors discipline

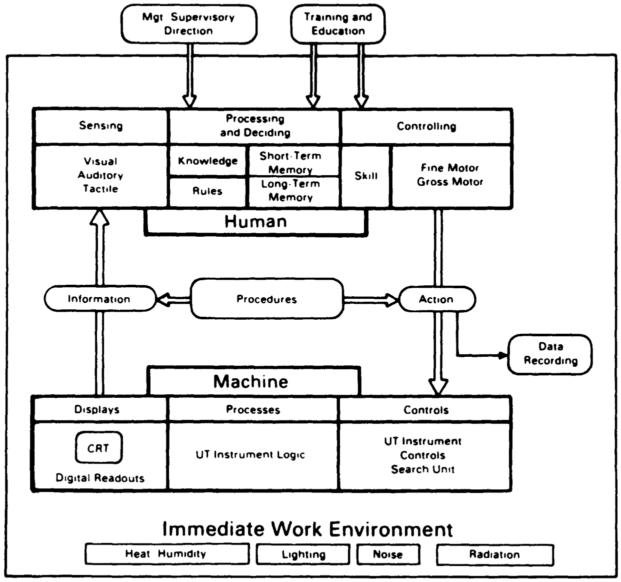

Fig. 2. Man-machine model of NDT system

The ROC analysis method was derived from classical signal detection theory and has been widely applied in other fields [8]. As with most measurement techniques, the ROC method requires a large number of test specimens to provide a sufficient data base. However, this deficiency can be ameliorated by using the rating scale technique to reduce the number of specimens needed. Examples of the ROC analysis method are included under the discussion of mini-round robin results.

FACTORS THAT AFFECT NDT RELIABILITY

Factors affecting the reliability of NDT include training, experience, procedures, equipment features, working environment, and the psychological pressures that are prevalent during all inspection operations. While the impact of human factors on NDT reliability has now been recognized, the variables influencing the performance of the NDT process have not yet been systematically studied.

Based on our work to date, the training variables appear to exert the greatest influence on technician performance (positive, negative, or both); hence, the training variables merit primary consideration. Research findings strongly suggest that the type of training most likely to improve performance is "hands-on" training using actual component specimens (prompt cueing and feedback should also be provided to the student). Significantly, there are no requirements for periodic training in the NDT field, and very little occurs except when upgrading from Level I to Level II. A parametric pilot study is needed to assess the value of periodic (e.g., annual) training (both hands-on and classroom).

The limited research available on procedural variables shows that better procedures could improve NDT performance, and this aspect definitely warrants more study. Time-on-task and the length of the work day also appear to influence performance. It is important to determine whether, and to what extent, a technician's performance changes during the course of a shift.

Many of the task variables can also be adjusted to improve NDT system performance. Some equipment appears easier to use because of differences in the amount of signal processing that is done and because of a sharper display image. The layout and labeling of controls and displays should be evaluated, as well as signal processing and screen image clarity effects.

Environmental variables can also impact human performance. Environmental variables include heat, humidity, lighting, protective clothing, cramped and awkward working conditions, fatigue, and radiation exposure. Since adverse environmental conditions have a negative impact on performance, systematic study of the environmental variables to measure the magnitudes of their effect on performance is warranted.

MINI-ROUND ROBIN RESULTS

The human factors concepts described herein were applied during a mini-round robin (MRR) study conducted in mid-1985 [9]. The overall approach used during this human factors evaluation is depicted in Fig. 3. Our primary goals were to: 1) acquire and evaluate data on the performance-shaping factors that affect UT/ISI reliability, and 2) evaluate the ROC analysis method using UT data acquired during the MRR. A written questionnaire and a significant incident interview were used to assess the technician's perceptions concerning various performance-shaping factors, and to analyze the relationship between personal characteristics and performance. The ROC analysis method was used to evaluate UT/ISI performance, and a cursory human engineering evaluation of the equipment was conducted.

A summary of the responses provided by the MRR participants on the questionnaire and during the interview is as follows:

- The nuclear industry's qualification process did not measure ability to detect IGSCC because the technique used to pass the qualification test was different than the techniques used during actual plant inspections.
- The utilities will probably verify crack calls, but will probably not verify non-crack calls.
- Better written procedures would improve NDT performance.
- The most valuable aspects of the industrial training received were: 1) an opportunity to practice on specimens with actual cracks, and 2) an opportunity to compare techniques with other personnel.
- The least valuable aspects of such training were: 1) the classroom lectures, and 2) the lack of detailed feedback on individual performance.
- The current certification levels of NDT personnel are not related to ability.
- It is expected that different UT technicians will often report different results.

The ROC analysis method was used to analyze individual technician performance in detecting IGSCC using specimens of austenitic stainless steel piping. The cracks were located on either the near or far side of a circumferential weld, and varied from about 0.635 cm (1/4 inch) to

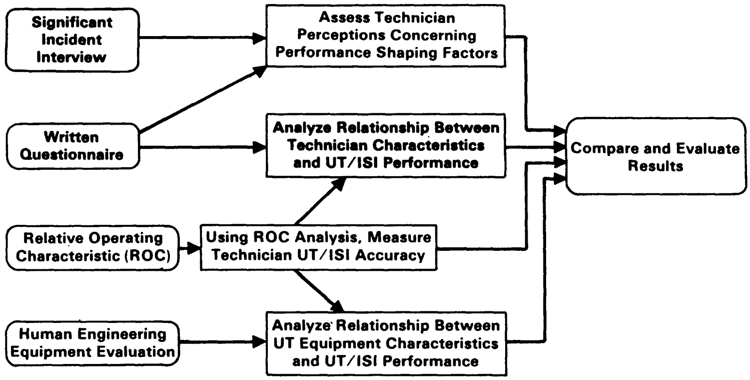

Fig. 3. Flow diagram of human factors evaluation conducted during mini-round robin study

greater than 15.24 cm (6 inches) in length. Of the twelve technicians that participated in the mini-round robin, eight were certified as Level III and four were certified as Level II. All twelve had successfully qualified for IGSCC detection, and at least six were also qualified for IGSCC sizing. Additional details on the experimental design and the conduct of the MRR are available in References 9 and 10.

Fig. 4 shows the average performance for near-side and far-side crack detection. The light diagonal line corresponds to performance based on pure chance (e.g., by flipping a coin), and the box in the upper-left corner is the "pass" criteria used during a "requalification" process that was implemented in late 1985. This figure shows that the average performance for near-side cracks did not meet the current requalification criteria, and that the average performance for far-side cracks was only slightly better than pure chance.

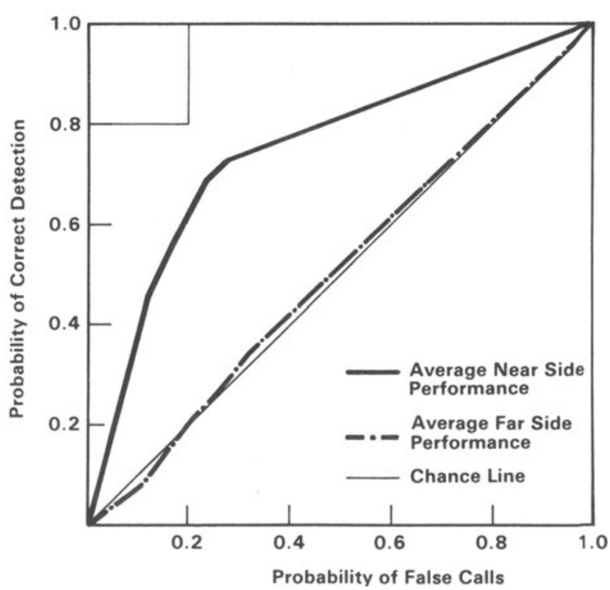

Fig. 4. ROC curves showing average performance for near- and far-side crack detection during mini-round robin study

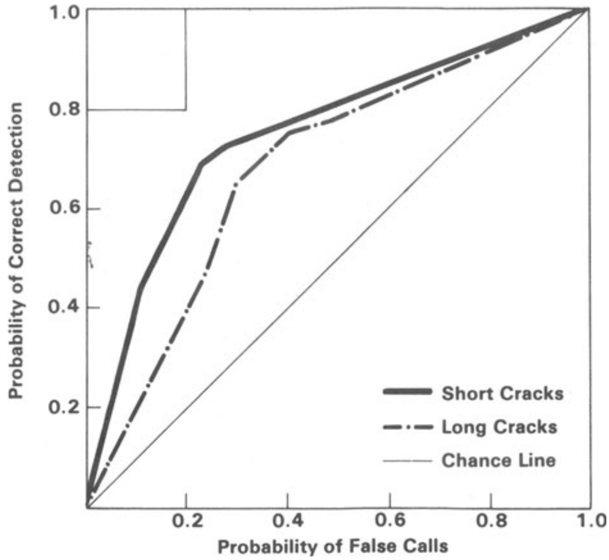

Fig. 5. ROC curves showing average performance for short (<2") and long (>3") near-side cracks during mini-round robin study

Fig. 5 shows the average performance for short and long near-side cracks. Note that the capability to detect short cracks was slightly better than that for long cracks. This result is in direct conflict with the usual assumption that long cracks will be detected more reliably than short cracks.

Fig. 6 dramatically illustrates the differences in individual performance that were observed throughout the MRR. Note that the curves for the three best performers fell within the requalification criteria now

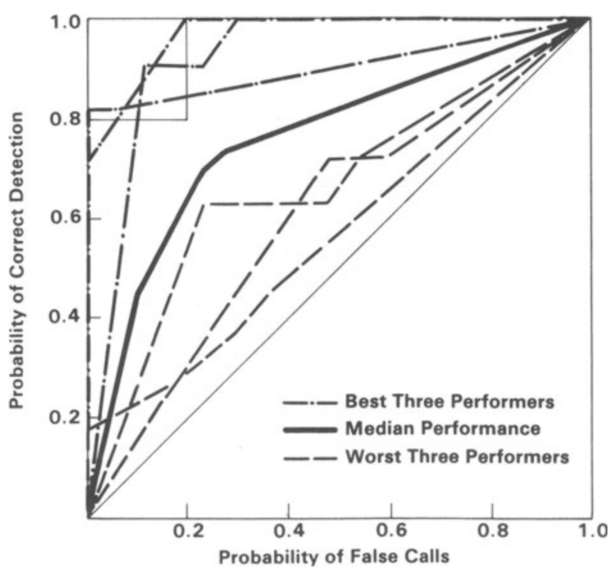

Fig. 6. ROC curves showing the three-best, median, and the three-worst performers for near-side cracks during mini-round robin study

in effect, but that the median performance for this group of technicians fell outside these criteria. Note also that the curves for the three worst performers were far below the requalification criteria, and that the performance of one individual was only slightly above the chance line for near-side cracks. Incidently, that individual's performance curve actually fell below the chance line for far-side cracks.

The reliability problems illustrated in Fig. 6 should be recognized whenever NDT data are being evaluated. Perhaps more importantly, this figure shows both the need and the potential for improving the reliability of the UT/ISI process. It also shows why additional research to evaluate the applicability of the available human factors principles is so urgently needed.

CONCLUSIONS

The following conclusions are based on this limited feasibility study.

- The current UT/ISI systems have a higher intrinsic capability for reliable flaw detection than is achieved in actual practice.
- It is useful to consider the NDT process as a semi-automatic man-machine system consisting of a human operator, plus the hardware, written procedures, and environment.
- The Relative Operating Characteristic (ROC) analysis method offers a unique capability to measure flaw discrimination performance that is unaffected by the various factors that influence the decision criteria. This capability is not available when using traditional measurement methods.
- The ROC analysis method may be used in conjunction with the rating scale technique to provide an objective, accurate, and cost effective measurement of NDT system performance.
- This project has shown that there is significant potential for improving the reliability of the NDT process by applying the concepts and principles that exist within the human factors engineering discipline.

This study has indicated that no single performance-shaping factor is responsible for the wide performance variations (i.e., UT/ISI unreliability) that have been observed. Hence, improvement of any single performance-shaping factor will not have sufficient impact to correct this problem. However, we are convinced that if the key human factors variables were all simultaneously adjusted in a positive direction, the resulting synergistic improvements in actual NDT performance would be significant.

This study has confirmed the need to investigate the human reliability aspects of the UT/ISI process and quantify the potential for improvement. A comprehensive human factors study should yield results that can be applied to improve the reliability of the other NDT methods that are used throughout industry.

ACKNOWLEDGMENTS

The support and encouragement provided by many individuals at the Pacific Northwest Laboratory (PNL) and the Nuclear Regulatory Commission

(NRC) during the course of this project is gratefully acknowledged. This work was supported by the U.S. Nuclear Regulatory Commission under Contract DC-AC06-76RLO 1830, NRC FIN B2883 and B2983.

REFERENCES

1. W. H. Lewis, W. H. Sproat, B. D. Dodd, and J. M. Hamilton, Reliability of Nondestructive Inspections--Final Report, SA-ALC/MME-76-6-38-1, Lockheed-Georgia Co., Marietta, Georgia (December 1978)

2. S. R. Doctor, F. L. Becker, and G. P. Selby, "Effectiveness and Reliability of U.S. Inservice Inspection Techniques," Proceedings of the Inst. of Mech. Eng. Conf., London, England, C150, 291-294 (1982)

3. W. D. Rummel, S. J. Mullen, B. K. Christner, F. B. Ross, and R. E. Muthart, Reliability of Nondestructive Inspection (NDI) on Aircraft Engine Components, Report No. SA-ALC/MM8151, Martin Marietta Corp., Denver, Colorado (1984)

4. J. C. Spanner, R. V. Badalamente, W. L. Rankin, and T. J. Triggs, Human Reliability Impact on Inservice Inspection, Vol 1. - Phase I Summary Report, NUREG/CR-4436, Pacific Northwest Laboratory, Richland, Washington (1986)

5. T. J. Triggs, W. L. Rankin, R. V. Badalamente, and J. C. Spanner, Human Reliability Impact on Inservice Inspection, Vol. 2 - Review and Analysis of Human Performance in Nondestructive Testing (Emphasizing Ultrasonics), NUREG/CR-4436, Pacific Northwest Laboratory, Richland, Washington (1986).

6. D. H. Harris, "Human Factors Success Stories," in Proceedings of the Human Factors Society Annual Meeting, 28, 1-5 (1982)

7. E. T. Klemmer and D. J. Dooling, "Making Human Factors Effective in Industry: Examples from Business Telephone Systems," in Applied Ergonomics, 14, 279-283 (1983)

8. J. A. Swets, "Assessment of NDT Systems--Parts 1 and 2," in Materials Evaluation, 41, 1294-1303 (1983)

9. W. A. Wheeler, W. L. Rankin, J. C. Spanner, R. V. Badalamente, and T. T. Taylor, Human Factors Study Conducted in Conjunction with a Mini-Round Robin Assessment of Ultrasonic Technician Performance, NUREG/CR-4600, Pacific Northwest Laboratory, Richland, Washington (1986)

10. T. T. Taylor, D. J. Bates, and P. G. Heasler, "An Evaluation of Current Capabilities in Characterization of Intergranular Stress Corrosion Cracking," in Proceedings of 8th International Conference on NDE in the Nuclear Industry, Orlando, Florida, American Society for Metals (1986)